国外古籍数字化标准研究进展及其启示

彭媛媛*

（中国科学院文献情报中心，北京，100000）

摘要：[目的/意义] 为我国古籍数字化标准化工作提供参考建议。[方法/过程] 文章对欧洲、美洲、大洋洲、亚洲地区的 7 个国家古籍数字化标准现状进行调查研究。[结果/结论] 梳理并总结出这些国家古籍数字化标准体系规范的主要内容及其应用特征。

关键词：古籍数字化 标准体系 研究现状 实践进展

Research Progress and Inspiration of Digital Standard of Foreign Ancient Books

Peng Yuanyuan

（National Science Library，Chinese Academy of Sciences，Beijing，100000，China）

Abstract：In the paper, according to Europe, America, Oceania, Asia the seven national countries’ ancient books digitized standard status investigation and study, summarized the main content of national ancient books digitization standards system specification and its application feature, in order to provide the reference suggestions for the Chinese ancient book digitization standardization work.

Key Words：Digitization of ancient books; Standard Architecture; Status quo; Practice progress

1. 前言

古籍是古代书籍的简称，这里的“古”是相对于“今”而言的，一般认为未采用现代印刷技术而印制的书籍皆可称之为古籍。我国《古籍著录规则》（GB/T3792.7-2008）中将古籍定义为“古籍主要是指 1911 年以前(含 1911 年)在中国书写或印刷的书籍。”[1]。这一定义指明了古籍的时间、地点和形式的三大关键要素。在国外，古籍的概念更为宽泛，不仅仅指古代书籍(Ancient book)，还包括善本（Rare book)、手稿（Manuscript)、拓片（Rubbing）等。古籍数字化是利用现代计算机技术手段将语言文字或图形符号转化为书目数据库或全文数据库的过程，用以利用和保护古籍[2]。

国外古籍数字化工作十分重视，许多国家开展了古籍数字化项目的建设。这些项目有些是针对本国善本、手稿等的数字化项目，各大洲的许多国家在古籍数字化项目实践的基础上制定了一些标准规范用以指导古籍数字化工作的进行，这些标准涉及古籍数字化的过程、技术和管理等方面。本文通过对国外 7 个国家的古籍数字化标准规范的主要内容及其应用特征进行研究整理，以期为我国今后古籍数字化标准工作的开展提供可行性参考范例。

2. 国外古籍数字化标准研究现状

在数字化过程中，内容的数字化和服务提供的多元化会使标准的建设难以达到一致性。古籍数字化标准的作用是什么。国外相关研究阐释了古籍数字化标准的相关理论问题，相关研究内容较为集中。

2.1 关于古籍数字化标准统一的研究

关于古籍数字化标准，目前在国外的古籍数字化研究中也没有统一的界定。在 Tanyildiz Ahmet 的《SüRELi YAYINLAR BAGLAMINDA BAZI CEVRiYAZI SORUNLARI》一文中讨论到虽然目前国际上有很多关于古籍数字化的研究，但是仍然没有一个统一的数字化标准。目前的研究成果数量有限，并且时间大多是在 2000 年之后。关于数字化标准中的协议和参数问题，他也提出了自己的意见。

2.2 关于数字化标准应用的研究

Jon Zabala-Vázquez 在《Utilidad y Vigencia de las Tipobibliografías y su Ausencia en el Mundo Digital》一文中分析了各种古籍数字化标准的重要性和有效性。文章基于 Tipobibliografía Espanola 项目，但是由于该项目发表的格式并不是以数字的格式，因此主要的古籍数字化自动的标准和描述性规则的意义不是很大。此外，作者还对比分析了 AACR2，ISBD(A),MARC 和 DCMI 的各个元素在描述上的标准，通过研究发现这些标准几乎适用于所有在线发布的文件。

3. 现行古籍数字化标准体系规范

本研究于 2016 年 1 月—2016 年 10 月，对世界主要国家的国家图书馆、政府机构网站、国际图书馆协会联合会（IFLA）等标准化机构、各大索引数据库等进行调查，了解各国古籍数字化的发展现状及其标准体系构建情况，并对各国古籍数字化标准及规范指南进行检索、归纳和整理，得出如下内容。

通过网站的调查研究，共检索到 50 余项相关规范，对检索到的文献进行二次筛选，最终得出 27 条古籍数字化的标准规范。国外许多国家虽然不同程度的对本国保存的古籍进行了数字化工作，但数字化程度、应用标准、数字化具体环节等方面存在许多差异。从表 1 的数据来看，尚未有一个国家形成完整的数字化标准工作流程或建立出标准体系框架。大部分国家是通过国际合作或部分开发的策略，逐步完善数字化工作。以下，对本文整理的各国标准规范及其应用做简要概述。

3.1 美洲国家标准规范及其应用

美国古籍数字化标准工作最为成熟，检索到 14 项文件，古籍数字化标准有 6 项，其中 1 项是由 IFLA 国际图书馆联合会发布的国际标准。其内容涉及善本与手稿的数字化标准、图像标准、编目标准、元数据标准、目录描述标准、保存标准、成像描述、缩微技术、文档要求等多个方面[3]。

加拿大古籍数字化的兴起是随着互联网的发展逐渐开始的。从 2002 年起，加拿大政府开始参与到图书馆数字化工作的进程中，国家图书档案馆也开启了 Amicus 的数字图书馆项目，对国家收藏的古籍版本进行数字化处理。在资料调查中发现，加拿大收藏有大量的手稿（如：希伯来手稿），一些专业图书馆为妥善管理专业原始手稿而建立专门的网站进行维护[4]。目前，能够获取的有关加拿大古籍数字化标准的共有 4 个。其中，包括对古老文献的保存与修复标准、文献著录编目的国际标准和加拿大国家图书档案馆的相关政策。

3.2 欧洲国家标准规范及其应用

英国古老文献数字化工作主要由英国国家图书馆负责，政府部门和一些企业机构给予资金方面的支持。现查找到的关于英国古籍数字化标准主要有 5 个参考资料。从标准规范的内容，英国古籍数字化标准工作主要应用到标准化管理、元数据、数据加工、文档文件存储等标准规范。英国现存有大量中国古籍，为了方便阅读和学习，英国与其他国家联合组建 IDP 项目、大英图书馆也独立建立了检索数据库，并应用数据库检索规范作为指导。

法国国家图书馆、法兰西比利时皇家藏书馆等机构中保存了大量法国王朝、皇室家族的手稿。能够获取全文标准规范资料仅有 1 项，Attractive guidelines for users EUROPEANA REGIA CIP-ICT-PSP 2009-250560，是针对用户使用方面的标准。法国对手稿进行数字化的过程中，注重元数据的规范和转化，制定较为详细的元数据标准规范。其次，为了完善对数字化手稿的管理，制定了相关的管理标准。满足用户检索的需要，有些图书馆建立手稿数据库，根据国际标准并结合法国实际情况，实行相应的数据库检索规范。

德国文献的数字化工作现处于较为分散的状态。由于历史原因，德国没有所属的国家图书馆，许多标准工作都是由科研机构完成，科研机构的能力和资金有限，标准工作处于起步阶段。能够获取全文的德国古籍数字化标准的文件资料仅有 1 个，DFG Practical Guidelines on Digitisation。该标准从数字化标准规范内容划分，主要包括元数据标准、文档存储标准。在数据加工方面，德国采用图像处理技术，将图像转换成固定的格式，可以用于网站访问和数据库检索使用。德国在数字化方面比较注重数字化成果的兼容并用、交流共享。

3.3 澳大利亚标准规范及其应用

澳大利亚国家图书馆从 1960 开始致力于将本国所有的文献资料全面收集、系统整理。其中包括古老、原生的图书、报纸、地图、手稿等资料。为了确保文献的长期保存，并能够在官网上进行虚拟访问，澳大利亚开始投资数字化图书馆项目。目前可获取古籍数字化标准共有 3 项。澳大利亚国家图书馆网站为“馆藏资源数字化”成立了专题区，在对古老文献数字化过程中，注重古籍版本的选择。大部分手稿通过扫描技术转化成静态图像储存，图书资料则直接数字化的情况较多。在数字化过程中，主要运用到文档存储标准、数据加工标准、资源的分类标准以及元数据标准规范等。

3.4 日本标准规范及其应用

日本与中国同属于东亚地区，都有着大量的历史文献。日本非常注重对文献的储藏和保管，制定了一些关于古老文献的保存和数字化标准，所获取的标准共 7 项。由于存在访问权限的问题，部分标准不能获取到全文，但可以从相关网站的描述中得知其主要包括著录标准和元数据标准。日本国立国会图书馆中存有大量善本手稿，因而建立了善本手稿的资料库，作为归纳整理的工具，也为用户的使用检索提供了方便。

4. 古籍数字化标准类别及其特点

为了进一步分析研究各国古籍数字化标准规范的内容、具体应用范围与使用情况，以及这些标准文件对今后古籍数字化标准工作的重要意义，本文运用内容分析法对以获取到的标准文件进行词频切分和内容重组，对不同关键词出现的频次进行统计整理，并根据其相应文件所属的标准大类和具体类别进行了细致划分，得出各项标准规范。

4.1 管理标准

古籍善本即记录着历史，也传承着文明。因此，许多国家对古籍、善本、手稿、特藏等文献资料都提出了适应于各国需要的管理标准；国际上，也提出了相应的国际型的标准。IFLA（国际图书馆协会联合会）于 2014 年 9 月颁布了由国际图联善本与手稿组撰稿而成的《馆藏善本与手稿数字化计划指南》[5]，指南中提及“在可发现性和使用方面，善本和特殊馆藏的数字化，比一般馆藏更有意义。一旦这些馆藏易于获取，它们就将成为核心资源”。国际图联制定相关指南的另一个目的在于，对孤本、善本和非印刷形式的馆藏进行数字化时，需特别加以考虑，制定特定的操作程序，旨在为善本、特藏的数字化提供解决方案，为图书馆

的馆藏存储工作提供参考意见；英国、澳大利亚、德国、日本等也出台了相应的古籍标准化管理规则。

4.2 数据标准

4.2.1 文档存储标准

英国现有的数字化项目、方案多采用SGML（标准通用标记语言）或采用XML（可扩展标记语言）的文档编码技术。其目的是将图像与单一主件或整体馆藏中允许访问的结构性元素相结合。文件编码的数据也可用在储存基础数据库系统中，并转化成标准化表示形式用以交换使用。但是，SGML存在对许多应用程序软件都不支持的问题；而XML已经开始逐渐接替作为最常用的标记语言[6]。现阶段的古籍数字化工作是基于SGML著名编码规则的TEI（文本编码倡议），该公司开发的DTD（文档类型定义）在人文领域对各个文本进行编码。编码档案说明倡议（EAD）开发了关于DTD编码检索工具，用来编码整个馆藏文献或其他对象。英国数字化古籍文献的文件格式大多通用、兼容。根据统一的文档存储标准，英国对完成数字化处理的文本文件进行标准化，以规范文档的存储格式。PDF存储格式的文件居多；少量文件可以通过Word文档形式读取。基于Web网页浏览版的文件，以相互兼容的HTML、XML、SGML核心语言编码，读者可以进行全面的浏览、检索与获取。

国外对中国古籍文献数字化的存储过程中，会涉及字符处理、技术规范等问题。其中，字符处理的标准，主要体现在古文献中存在的生僻字词，不为人所熟知，也甚少使用。尤其是涉及到医学、天文、地理方面的字词及其表达的含义，非系统化的描述和规范，为实施相应数字化工作带来较大的困难。许多国家规定，数字化之前对文献中包含的字符和相关文字术语进行标准化和规范化处理，根据现有的字符集标准，进一步规范处理，形成通用的字符集。在技术标准方面，由于大量古籍版本形式不同，纸张性质差异较大。为实现最适宜的数字化工作，在版本的选择和操作上，应用技术标准加以限定和约束。将其转化为数字化格式的文件，以达到国际通用、资源共享的目的。应用现行文件格式编码标准，对数字化文献格式进行控制，尽可能转换为PDF、HTML文件格式[7]。

4.2.2 元数据标准

在存储手稿资源方面，为了实现更为妥善的保护，许多国家都会选择限定主文件的文件格式，依据既定的元数据标准进行保存，同时区分不同的存储介质。标准规范中将元数据的类别主要分为三种，即书目（或描述）型元数据、图像（或技术）型元数据、结构型元数据三种。书目元数据描述了被数字化的实体对象，

并包括知识内容方面的信息。所有被挑选出来要进行数字化的材料在数字化之前都要有某种书目元数据。在完成馆藏文献数字化后，应把指向数字版本的链接追加到相关书目记录上，该链接还要从该数字版本指回书目记录。图像（或技术）型元数据通常是由相机或扫描仪自动获得，并出现在文件的标头。主要包括：以像素为单位的长度和宽度、采样、压缩方法、分辨率、以字节为单位的大小等内容；如果图像是加工处理过的，与之相关的信息也要包含其中并且能够传递给用户。结构型元数据用于研究较为复杂的对象，比如中世纪的手稿、档案、信件或照片相册，研究者需要能够从一张张数字页面或图像中重新创建（或重组）实体对象。正在对中世纪手稿和其他类似复杂对象进行数字化的图书馆，应特别注意要进行优秀的校对，并提供其他形式的结构型元数据。至少，要保证研究者能够确定页面或图像的原始顺序。

许多图书馆需要对数字文件进行存取管理时，会使用到第四种元数据——管理元数据，加以协助。管理元数据可能包括图片的拍摄者、图像或者原件的所有者、版权信息以及使用额度。管理元数据还可能包括如下信息：数字化过程中没有进行数字化的馆藏项目及其原因和形成的决议，还有常规的图像处理过程（整个馆藏的处理过程是一致的）的说明。

4.3 描述标准

国际为了方便加工和组织馆藏信息资源，提出了 FRBR（书目记录的功能需求）、FRAD（规范数据的功能需求）、FRSAD（主题规范数据的功能需求）三个概念模型。在此基础上，为了实现馆藏信息资源的著录，提出了 ISBD 国际书目著录标准。为了满足数字化格式的编码需求，进而制定了 UNIMARC 编码格式。UNIMARC 由美国提出的 MARC 格式加以演变而成的，它的出现得到了许多国家的认可，各国也在此基础上形成适于各国需要的 MARC 字段。自 19997 年，“AACR 理论与未来发展”的国际会议中，首次出现 RDA（资源描述与检索）这一概念模型。RDA 是基于 FRBR 与 FRAD 两个概念模型演变而成的，其超越了过去的编目规则，提供关于数字资源编目的指南，更加强调帮助用户查找、标识、选择和获取所需信息。

索引标准各国最常用的是由美国国会图书馆提出的主题词(LCSH)编目标准。当 LCSH 发布实施以来，在国际上具有较高的使用率，各国的图书馆界也纷纷以此进行参考借鉴，对于有特殊需要，或者古籍文献具有较强的专门性时，会根据需要和习惯进行修改和调整。

4.4 工作标准

各国创建数字馆藏的过程都可以分解成不同的阶段或步骤。各图书馆的操作情况或许稍有变化，但基本流程是大致相同。具体数字化工作实施之前，准备并

检查所需的数字化原件以及相关材料，确定是否具有实体条件和现有的元数据资源。从文献库中提走所需要的数字化原件，对其移动和使用情况紧密跟踪，并记录下各时期的存在形态和存放情况。应对脆弱易损的原件，要事前进行评估预测，以尽量减少数字化过程中文献实体受损情况的发生。扫描技术人员应接受适当的培训，以确保能安全地扫描原件。同时，图书馆工作人员需要预估当前的书目元数据是否能满足读者的需求。书目元数据应该是在同一个系统中，支持用户的检索，并能够提供指向数字化替代品的链接，和由数字化产品中得到的返回链接。编目是创建馆藏的一个重要环节，应该通过相关协议、描述水平和细节，以及用什么语言做出判断。作为最低要求，要有与主目录语言相一致的描述。为了参与到国家或国际项目中，使更多的读者能获取这些数字馆藏，一般会增补翻译成其他语言，以满足广大用户的需求。

5. 启示

本文基于对国外古籍数字化现状的调查研究，应用内容分析法对国外古籍数字化过程中形成的标准体系框架进行梳理，以期为我国古籍数字化标准化建设提供参考。

首先，国外古籍数字化标准建设相对比较全面，值得借鉴参考。本文对美国、英国、法国、德国、加拿大、澳大利亚、日本等国家古籍数字化标准及规范指南进行检索、归纳和整理。国外许多国家，都有保存其国家古老文献的习惯；历史因素的影响，大多数国家同时存储有大量中国的古籍资源。国外许多国家虽然不同程度的对本国保存的古籍进行了数字化工作，但数字化程度、应用标准、数字化具体环节等方面存在许多差异。大部分国家是通过国际合作或是部分开发的策略，逐步完善数字化工作。

第二，古籍数字化应形成某种周期性质，即项目规划、数字化内容选择、数字化生产流程、数字资源保存以及数字资源发布等几个部分。古籍数字化资源作为数字资源的一种特殊形式，其周期管理包括将某古籍转换为数字资源并实现共享利用的全过程，涉及项目规划、数字化内容选择、数字化生产流程、数字资源保存以及数字资源发布这五个阶段。

第三，古籍数字化标准从本质上来说，是有意识地对古籍数字化整体工作的规范，包括横向、纵向两个方面，不仅要考虑到标准化工作的全面性，覆盖整个范围，还要考虑得到标准工作的系统性，对当前问题和潜在问题都进行评估与规范；在一定条件下对标准对象做出统一规定；考虑到多方面的要求，要图书馆等机构协同参与标准的制定与执行。古籍数字化标准的具体活动过程一般包括标准的制定、实施、监督和修订，要严格依照《标准化法》、《国家标准管理办法》等

法规实施。在进行标准体系设计时，应当遵守全面系统、协调一致、规范具体、效益优先和变动有序的五点原则。

第四，基于国外现已形成的标准体系，构建我国古籍数字化标准体系。本文对国外古籍数字化的标准体系做了重新整理与归纳，并对一些部分进行了详细描述，但其依旧有所缺失和不足。从现有的标准体系中可以得出，我国的标准化体系应该从古籍数字化项目规划、古籍版本选择、古籍数字资源生产、古籍数字资源保存、古籍数字资源发布五个模块，分别面向古籍数字化的各个环节，能够做到全面系统地指导与规范古籍数字化工作

最后，我国古籍数字化标准化体系实施的国家战略。在重点领域和优先事项方面提出古籍项目规划与版本选择的标准规范，形成数字资源发布、生产、保存的规范内容；建立协调共建的保障机制，设立政府的监督机构；促进国际化合作项目，形成标准化战略；利用国家政策进行调控，培养专业性人才，进行资金支持和制度支撑，力求从宏观的四个层次上得出有利于我国古籍数字化标准化体系实施和发展的国家战略。

参考文献：

[1]GB/T 3792.7-2008，古籍著录规则[S].

[2]屈菡.古籍数字化将走向规范化[N].中国文化报,2012-5-23(2).

[3]T.wei Karen. 北美古籍文献资源建设与数字化的调查 [EB/OL].[2015-6-2].http://wenku.baidu.com/link?url=v8wxCXOcCpvH2f2dWajccW7vbKboSy3cfWzj4MOjj5_rWSnB6yPLucdNnu1XUjHdnplhEjo-GbJAace6tPiHxh_vDSHlhH1QSWPcnagYTP_.

[4]About Canadiana.org. [EB/OL]. [2015-06-26]. http://www.canadiana.ca/en/about.

[5]IFLA 官方网站[EB/OL].[2016-2-11].http://www.ifla.org/.

[6]英国古籍数字化项目指南[EB/OL]. [2012-05-11]. http://www.bl.uk/aboutus/stratpolprog/digi/digitisation/index.html.

[7]Preservation Activities. [EB/OL]. [2015-06-26]. http://www.bac-lac.gc.ca/eng/about-us/preservation/Pages/preservation-activities.aspx#a

（通讯作者：彭媛媛 pengyuanyuan@mail.las.ac.cn）